Intelligence Community and Department of Defense Content Discovery & Retrieval Integrated Project Team (CDR IPT)

IC/DoD SOAP Encoding Interface Specification for CDR Retrieve v1.1

12 May 2011

REVISION/HISTORY

Doc Revision	Revised By	Revision Date	Revisions
0.1		26 February 2010	Initial draft for subgroup review.
1.0-20100312	CDR IPT	12 March 2010	Minor updates.
1.0-20100319	CDR IPT	19 March 2010	Minor updates.
1.0-20100327	CDR IPT	27 March 2010	Minor updates.
1.0-20100331	CDR IPT	31 March 2010	Minor updates.
1.0-20100331	Farley	29 April 2010	Tech Edits

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Service Overview

The **Retrieve** Component, as defined by the "Intelligence Community/Department of Defense (IC/DoD) Content Discovery and Retrieval (CDR) Specification Framework" [SF], is the primary mechanism for content consumers to access one or more specific content resources from content collections. This component provides a common interface and behavioral model for IC and DoD content collections, enabling consumers to retrieve and initiate delivery of content resources. Specifically, the Retrieve Component provides a means to retrieve the native content described at a higher level by the Search Component query results.

This specification defines requirements and provides guidelines for the realization of the CDR Retrieve Component as a web service using the SOAP messaging protocol, hereafter termed a **Retrieve** service in this document. It describes a **Retrieve** service's behavior, interface and other aspects in detail, providing enough information for **Retrieve** service providers and implementers to create CDR-compliant **Retrieve** services.

The **Retrieve** service exposes a Retrieve function. While the function is often used in concert with retrieving results of a Content Discovery search, it may be used in general to process any compliant retrieve instructions. As discussed in CDR Specification Framework, a **Retrieve** service's results are the content resource. In the context of **Retrieve**, the content resource generally refers to the entire underlying record.

1.1.1 Scope

The **Retrieve** Component as defined *supports* the retrieval for a specified resource in a Content Collection. In addition, the Retrieve component can only support returning a resource directly to the requestor. It cannot redirect output to a component other than the requestor. Furthermore, no special handling instructions (e.g., routing) may be specified.

1.1.2 Relationship to Other CDR Architecture Elements[1]

The CDR Architecture prescribes an abstract-to-concrete model for the development of architecture elements and guidance for content discovery and retrieval. Each layer or tier of the model is intended to provide key aspects of the overall guidance to achieve the goals and objectives for joint DoD/IC content discovery and retrieval. The following graphic, discussed in detail within the CDR Reference Architecture [RA], illustrates this model.

[1] *For a detailed description of each of the layers, please reference the CDR RA Section 1.*

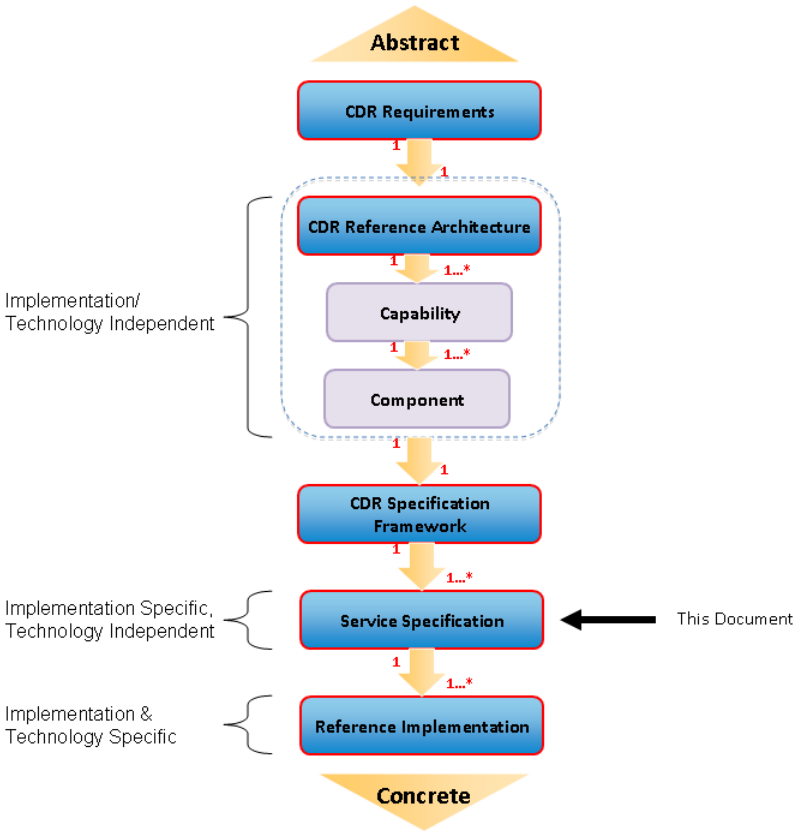

Figure 1 - CDR Architecture Model

As illustrated in Figure 1, the Specification Framework derives from the Reference Architecture (RA) and can describe behavior in terms of the capabilities, components, and usage patterns defined in the RA. The Specification Framework then expands on the details of information flows and the information conveyed in those flows to provide a consistent basis for multiple Service Specifications to provide consistent interfaces both in terms of the structure and the semantics of the exchanged information. Service Specifications, such as this one, provide implementation-specific guidance. More specifically this Retrieve Specification defines the specific guidance for implementing the CDR Retrieve Component as a SOAP-based service.

This specification covers the following aspects of a SOAP-based Retrieve Component:

- **Service Interface** defines the base SOAP constructs to expressing inputs, outputs, and faults
- **Implementation** provides additional implementation guidance beyond the behavior and interface guidance
- **Reference Documentation** provides references to other CDR and community artifacts (i.e., Service Security Reference Architecture)

1.2 Notational Convention

The key words "MUST," "MUST NOT," "REQUIRED," "SHALL," "SHALL NOT," "SHOULD," "SHOULD NOT," "RECOMMENDED," "MAY," and "OPTIONAL" in

this specification are to be interpreted as described in the IETF RFC 2119. When these words are not capitalized, they are meant in their natural-language sense.

When describing concrete XML schemas and example XML documents, this specification uses XPath as the notational convention. Each member of an XML schema is described using an XPath notation (e.g., /x:RootElement/x:ChildElement/@Attribute). The use of {any} indicates the presence of an element wildcard (<xs:any/>). The use of @{any} indicates the presence of an attribute wildcard (<xs:anyAttribute/>).

Examples in this text are distinguished by a black border. These are meant to be illustrative and only one way that the described syntax can be used.

1.3 Conformance

This specification defines an interface to a **Retrieve** service to which an implementation MUST conform. For an implementation to conform to this **Retrieve** specification, it MUST adhere to all mandatory aspects of the specification.

1.4 Namespaces

The following table only represents those XML namespaces that are directly leveraged in this document.

Table 1 – Referenced XML Namespaces

Prefix	URI	Description
soap	http://www.w3.org/2003/05/soap-envelope	W3C SOAP Version 1.2
cdrr	urn:cdr:1.0:soap:retrieve	CDR v1.0 Retrieve Service Specification for SOAP Implementations

2 Retrieve Service Interface

2.1 Retrieve Function

The **Retrieve** service specification is REQUIRED to function as described by the Content Discovery and Retrieval Specification Framework with any input, behavior, output, and fault condition extensions listed below.

Function	Input	Output	Fault
Retrieve	*cdrr:Identifier*	*xs:any*	Defined within CDR Framework

2.2 Input

In addition to the requirements imposed by the CDR Specification Framework, the following describes further input constraints on the SOAP Retrieve function:

- The `wsa:Action` element MUST be present in the SOAP Header and contain the value "*urn:cdr:1.0:soap:action:retrieve"*
- The *cdrr:Identifier* element MUST be present

The following examples illustrate two different mechanisms for specifying the *cdrr:Identifier* in the retrieve request message. In the first example, the *cdrr:Identifier* is inserted in the SOAP Header along with the *wsa:Action* and in the second, the *cdrr:Identifier* is present in the SOAP Body.

Example 1: Request Message with Identifier in the SOAP Header

```
<soap:Envelope>
  <soap:Header>
    …
    <wsa:Action>urn:cdr:1.0:soap:action:retrieve</wsa:Action>
    <cdrr:Identifier>urn:uuid:1225c695-cfb8-4ebb-aaaa-6fda344efa6a</cdrr:Identifier>
    …
  </soap:Header>
  <soap:Body/>
</soap:Envelope>
```

Example 2: Request Message with Identifier in the SOAP Body

```
<soap:Envelope>
  <soap:Header>
    …
    <wsa:Action>
      urn:cdr:1.0:soap:action:retrieve
    </wsa:Action>
    …
  </soap:Header>
  <soap:Body>
    <cdrr:Identifier>urn:uuid:1225c695-cfb8-4ebb-aaaa-6fda344efa6a</cdrr:Identifier>
  </soap:Body>
</soap:Envelope>
```

The following is a description of significant elements:

/wsa:Action
This REQUIRED element indicates the intent of the SOAP request
/cdrr:Identifier
This REQUIRED element contains the unique identifier for the requested resource, as shown in the example

2.3 Output

As imposed by the CDR Specification Framework, the following describes the output constraint on the **Retrieve** service Search function:

- The response information, including data or a collection of data MUST be present in the `/soap:Body`.

Binary data must be base64 encoded before inserting into the SOAP body. The following example illustrates the high level components of a response message (containing binary data) from a **Retrieve** service:

```
<soap:Envelope>
  <soap:Body>
        IkdpdmUgbWUgYSBsZXZlciBsb25nIGVub3VnaCBhbmQgYSBmdWxjcnVtI
        G9uIHdoaWNoIHRvIHBsYWNlIGl0LCBhbmQgSSBzaGFsbCBtb3ZlIHRoZS
        B3b3JsZC4iIC0gIEFyY2hpbWVkZXM=
  </soap:Body>
</soap:Envelope>
```

To further reduce client and server processing when encoding or decoding large amounts of data, the SOAP Message Transmission Optimization Mechanism (MTOM) SHOULD be leveraged.

2.4 Fault Conditions

An implementation of the **Retrieve** service MUST allow for the Fault Conditions defined in the CDR Specification Framework.

2.4.1 Identifier Execution Fault

During execution of the Retrieve Function, the Retrieve Component encounters an error retrieving the specified resource, and returns a “Identifier Execution” fault back to the Consumer.

2.4.2 Security Fault

During execution of the Search Function, one or more of the security components determines that the consumer is either not authenticated or not authorized to perform the search query. As a result the Search Component returns an appropriate security fault back to the Consumer as defined by the “Joint IC/DoD Security Reference Architecture. Also, for security considerations, see Section 3.2 of this document.

2.4.3 Fault Handling in SOAP

Different versions of SOAP may have different fault handling syntaxes. **Search** services MUST use the primary fault handling mechanism for the version of SOAP they support and to which the service is bound. In the following example, an Unsupported Identifier Type fault is returned using the SOAP 1.2 syntax:

```
<soap:Fault>
  <soap:Code>
    <soap:Value>soap:Sender</soap:Value>
  </soap:Code>
  <soap:Reason>
    <soap:Text>Identifier Execution Fault</soap:Text>
    <soap:Text>Service could not retrieve the specified resource</soap:Text>
  </soap:Reason>
</soap:Fault>
```

2.5 SOAP Version

Retrieve services may support SOAP 1.1 or SOAP 1.2. While SOAP 1.2 is RECOMMENDED, **Retrieve** service providers should support the version of SOAP that best meets their business objectives, referencing standard registries such as DISR[2] and ICSR[3], as appropriate, but MAY support both versions, if desired.

2.6 WSDL

Web Service Description Language (WSDL) documents are provided as supplements to this service specification that provide the necessary WSDL bindings for the **Retrieve** Service.

2 *DISR: DoD IT Standards Registry*

3 *ICSR: Intelligence Community Standards Registry*

3 SOAP Retrieve Specification Usage

This section provides additional implementation guidance beyond the behavior and interface guidance provided in the previous sections.

3.1 Policy

This specification defines the technical requirements and guidelines for implementing a **Retrieve** service. Policy for **Retrieve** service implementations is described in auxiliary documents. See the Reference Documents section for a listing of relevant policy documents. Implementers MUST follow the guidance in those policy documents.

3.2 Security Considerations

Any resource may have associated policies for use, especially as applies to authentication and authorization. These policies may be asserted by both the resource owner and those responsible for governance and management of the enterprise. The implementation of policies related to security considerations SHOULD leverage the specific security components and interactions defined by the Joint IC/DoD Security Reference Architecture (SRA), and MUST be in compliance with requirements and guidance for security outcomes as specified in the SRA and its associated specifications. ***Retrieve*** implementations MUST follow the guidance in those documents.

4 Reference Documents

The documents in this section either provide the foundation for, or define extensions to, or include implementation guidance for the **Retrieve** service. They include additional specifications, including those provided as part of the greater CDR specification set, and guidance documents that communicate current policy or implementation details. Each document is assigned a reference identifier, which is cited when the document is referenced within this Retrieve Service Specification.

In some cases, documents have been referenced with a version and data of "future" in order to track the iterative development of some of the extensions.

4.1 Specifications

4.1.1 (CDR) Specifications

The following documents provide a foundation and guidance for the development of this Retrieve Specification document. Retrieve service implementers should have a thorough understanding of the concepts and guidance in these documents; this Search Specification represents a realization of the Retrieve Component defined therein

Ref.	Title	Version	Date
SF	IC/DoD Content Discovery and Retrieval Specification Framework	DRAFT 0.6.2	29 Jan 2010
RA	IC/DoD Content Discovery and Retrieval Reference Architecture	DRAFT 0.4	16 Dec 2009

4.1.2 Other Specifications

4.1.2.1 Security Specifications

Ref.	Title	Version	Date
S1	Joint IC/DoD Security Reference Architecture	1.0	25 Jul 2008

4.1.2.2 Service Discovery Specifications

Ref.	Title	Version	Date
D1	Joint IC/DoD Service Discovery Architecture	DRAFT 1.2	28 Sep 2007

4.2 Policy and Guidance

4.2.1 (CDR) Policy and Guidance

This specification primarily addresses the behavioral and interface aspects common to all CDR **Retrieve** service implementations. The following document provides additional requirements and expectations set by policy or as a result of

Ref.	Title	Version	Date
P1	IC/DoD Content Discovery and Retrieval Retrieve Service Policy for SOAP Implementations	Future	Future